Report on ANN for NC - ν_eCC (all) classification

- **Goal :** Use Artificial Neural Networks to distinguish NC from ν_e CC neutrino interactions as well as NC from all the rest on "event-by-event" basis using topological and physical characteristics of neutrino events derived from MC generated interactions:
  - CC ν_μ ν_e ν_τ
  - NC
- **Methods :**
  - Construct two Neural Networks : a) **ν_μ CC - All the rest**
    b) **All the rest - NC**
  - Construct one Neural Network : a) **NC - ν_e CC**

MC events used for training

- For every period we construct a separate set of ANN's (3) since every period has different target configuration. We present the results for periods 3 and 4.

- Used 5000 events for each period with the following characteristics :

52 % ν_μ - 41% ν_e - 7% ν_τ 25 % NC - 75% CC

22% prompt/prompt+nonpromt

	ST1	ST2	ST3	ST4
period 3	**29%**	**24%**	**25%**	**22%**
period 4	**29%**	**27%**	**28%**	**16%**

Input variables used for training

- The variables we used for NC - ν_e CC, (ν_μCC - All the rest), NC - All the rest classification are :

HITS Total number of DC hits
(Total number of MID hits in the Central tubes)

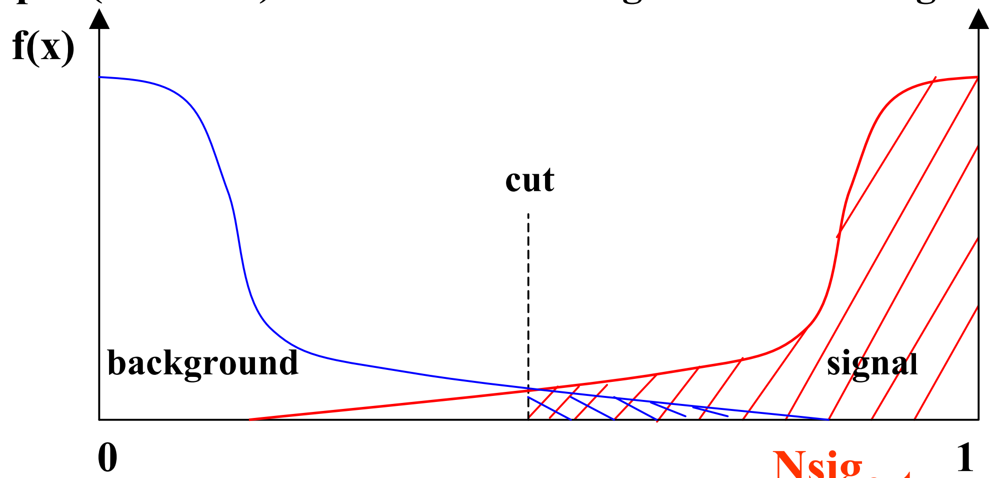
EMCAL Total energy deposition
Number of clusters
Average Cluster energy
Mean value of the Clusters angle from the vertex with respect to the z - axis
Standard deviation of the Clusters angle
Mean Absolute deviation of the of the Clusters angle
Higher Moments of the Clusters angle : a) Skewness b) Curtosis
(Percentage of tracks with E/P < 0.3 (Muons))

TRACKS Number of final tracks
Number of DC tracks
(Number of tracks that have more than 4 hits in the MID system (Muons))

OTHER Total Pulse Height in the SF system

Variables that characterize the ANN

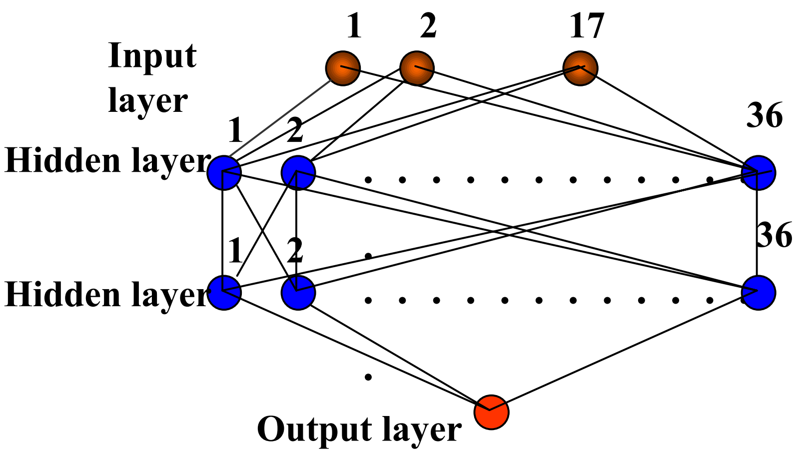
Network output (selection) function for “background ”and “signal” events



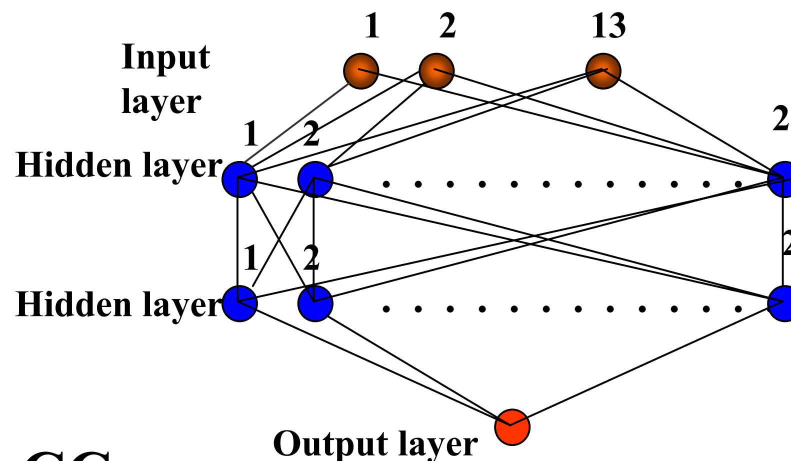
- **Signal Selection Efficiency :** $$\text{efficiency} = \frac{\text{Nsig}_{\text{cut}}}{\text{Nsig}}$$
  - Number of signal events above the cut / Total number of signal events
- **Signal Selection Purity :** $$\text{purity} = \frac{\text{Nsig}_{\text{cut}} / \text{Nsig}}{\text{Nsig}_{\text{cut}} / \text{Nsig} + (\text{Nback}_{\text{cut}} / \text{Nback}) * (\text{Nback} / \text{Nsig})}$$
  - Percentage of signal events above the cut / Percentage of signal events above the cut plus the Percentage of background events above the cut * (background to signal ratio)
- **Signal Selection Contamination :** $$\text{contamination} = \frac{\text{Nback}._{\text{cut}}}{\text{Nback}}$$
  - Number of background events above the cut / Total number of background events

Network's structures for period 3 and 4

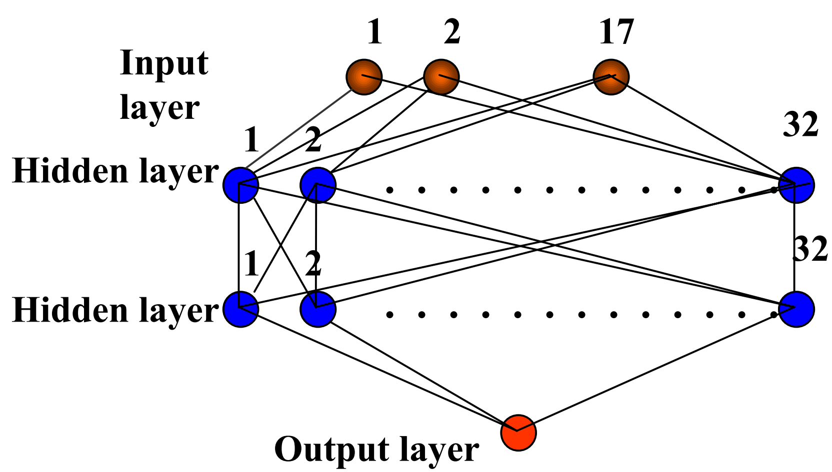
ν_μ CC - All the rest



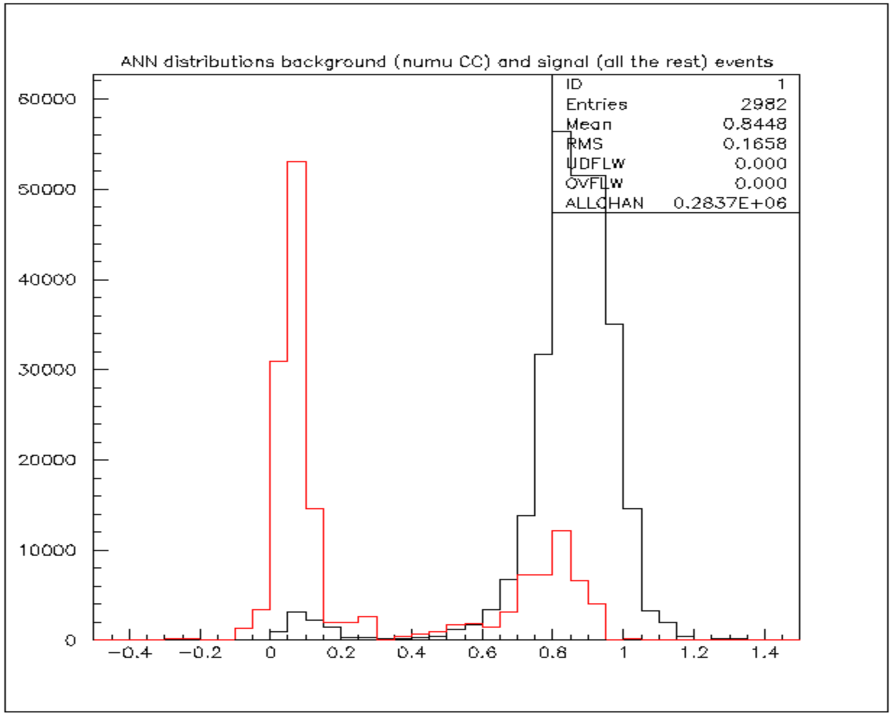
NC - All the rest

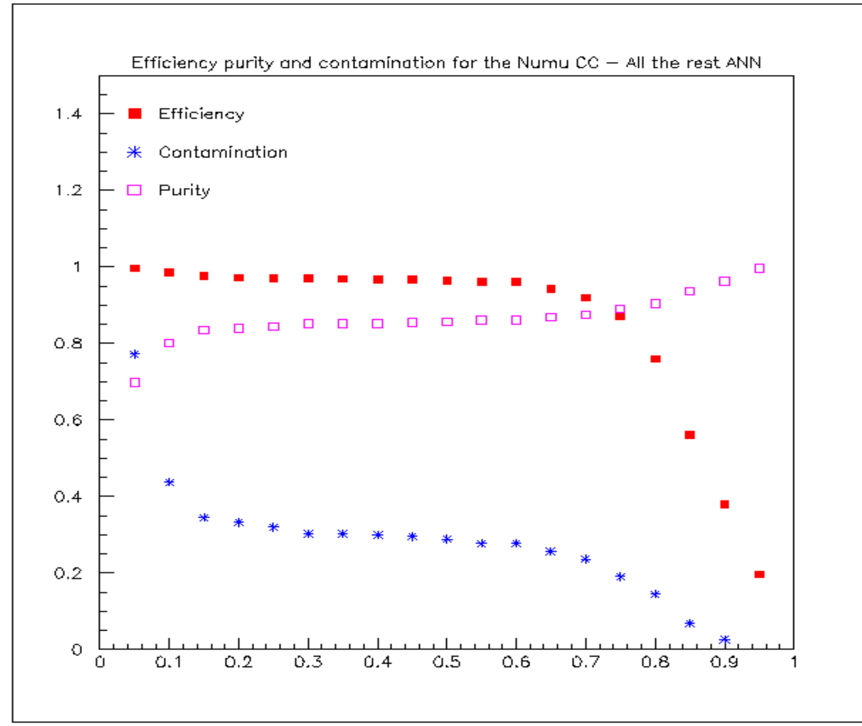


NC - ν_e CC



Results on ν_μCC - All the rest for period 3

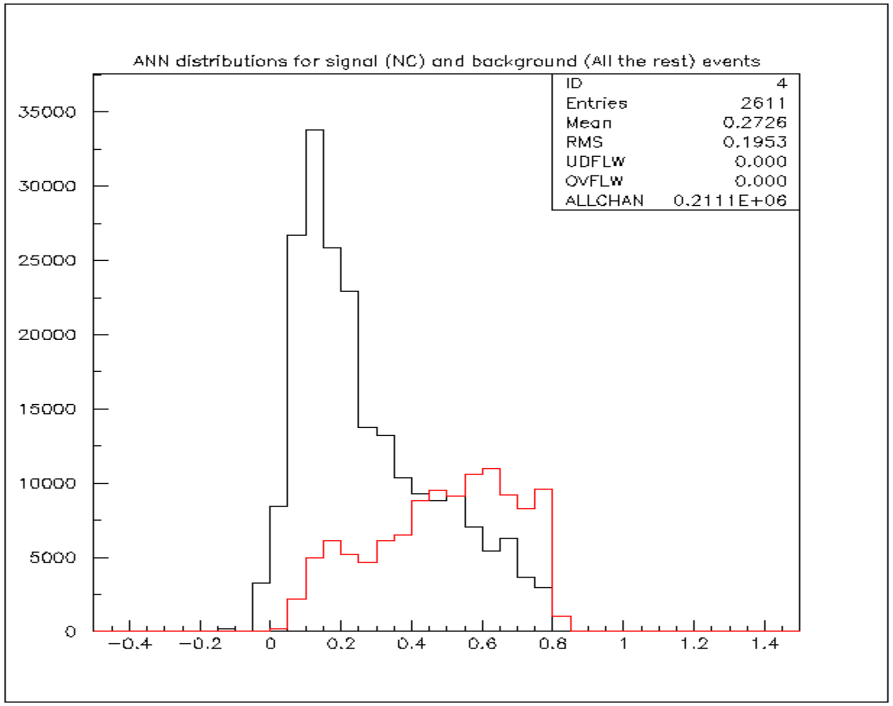


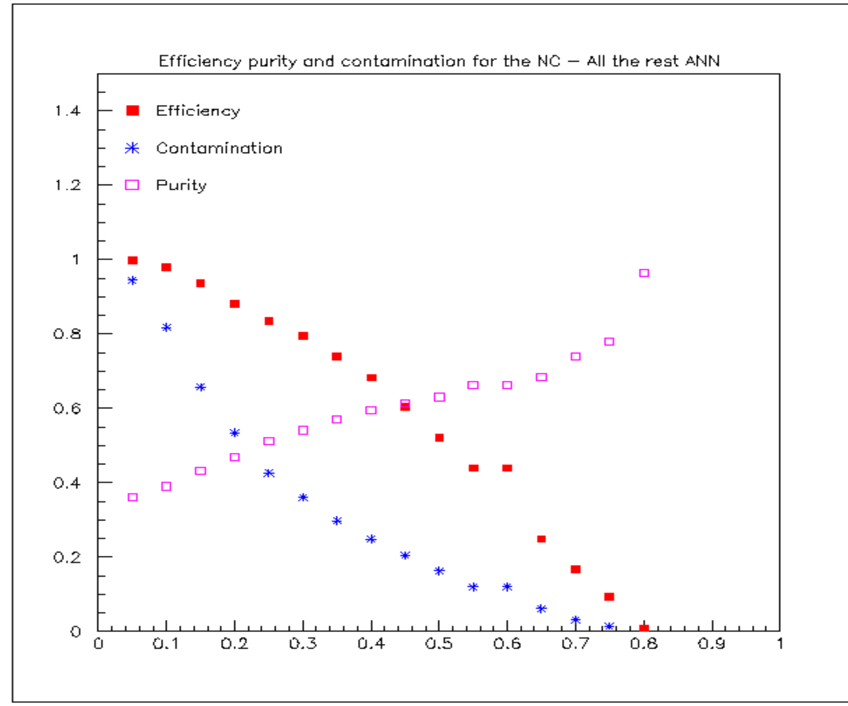


- If we consider a "cut" at 0.5 then we select signal (all events except ν_μ CC) with :

efficiency 96.5 % - purity 85.7 % - contamination 28.9 %

Results on NC - All the rest for period 3

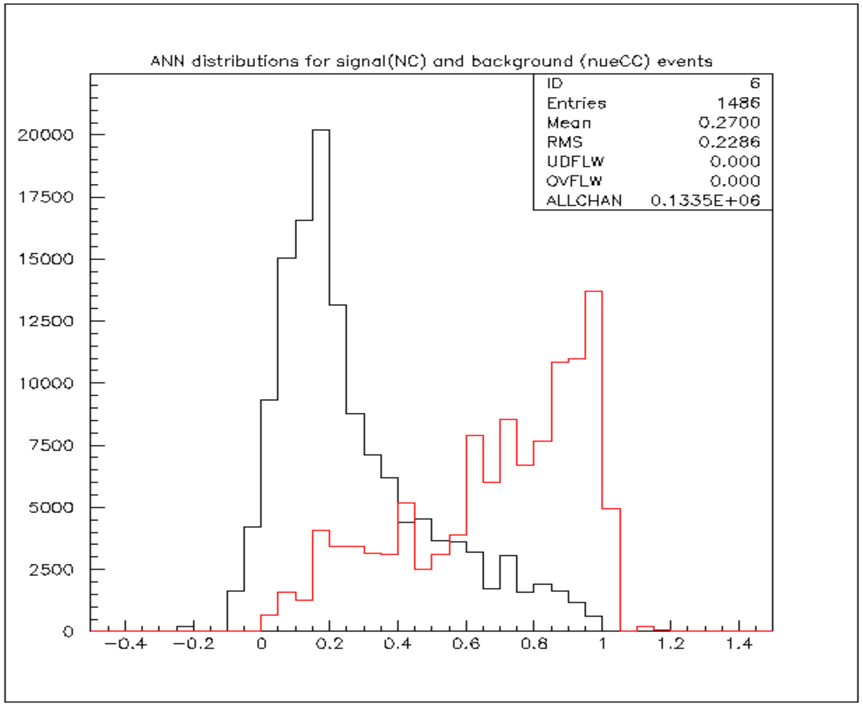


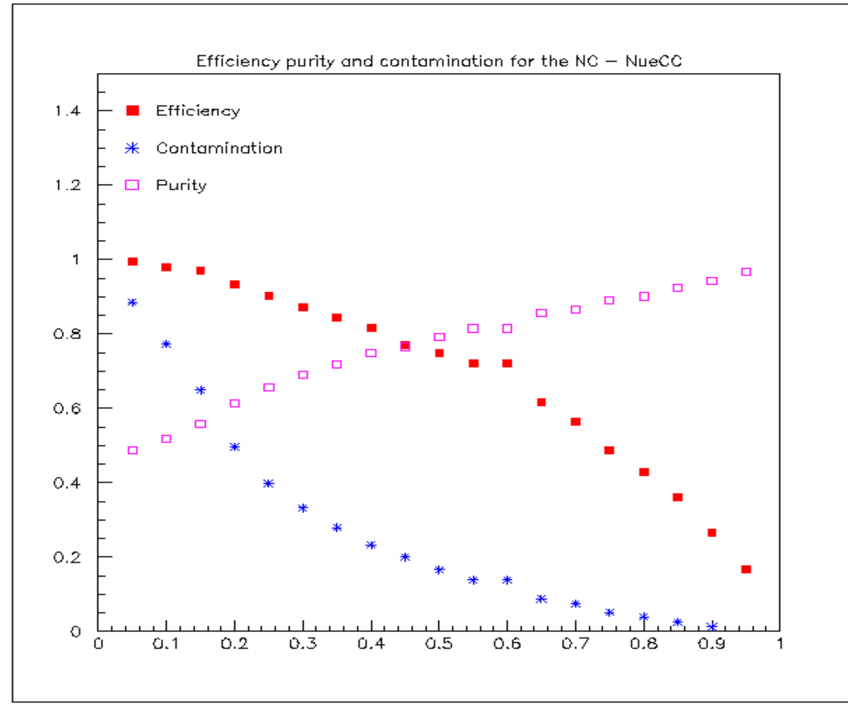


- If we consider a "cut" at 0.3 then we select signal (NC events) with :

efficiency 79.4 % - purity 54.1 % - contamination 36.0 %

Results on NC - ν_e CC for period 3

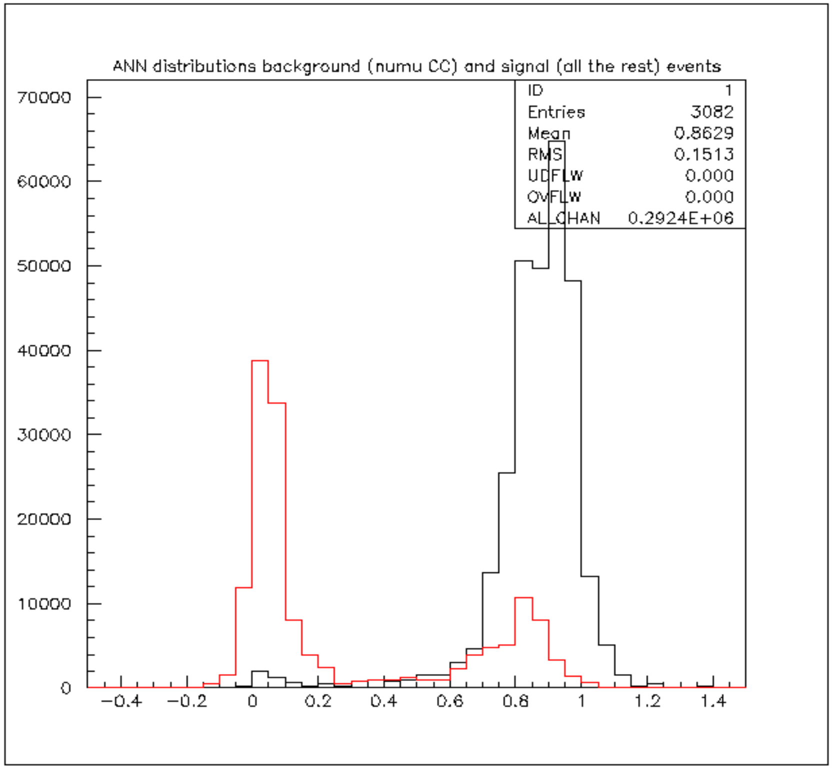


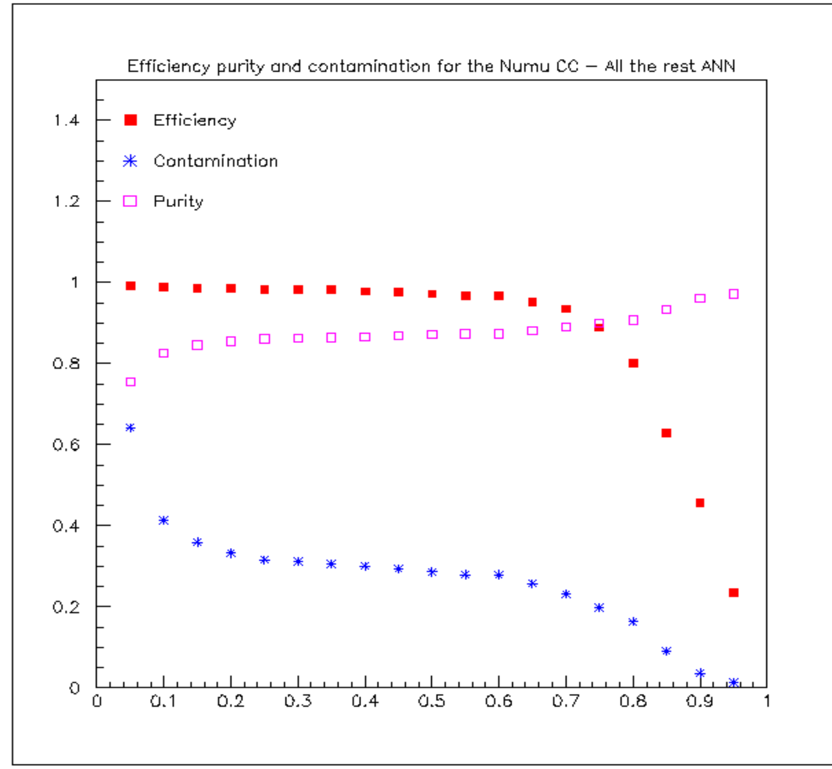


- If we consider a "cut" at 0.35 then we select signal (NC events) with :

efficiency 84.5 % - purity 71.9 % - contamination 27.9 %

Results on ν_μCC - All the rest for period 4

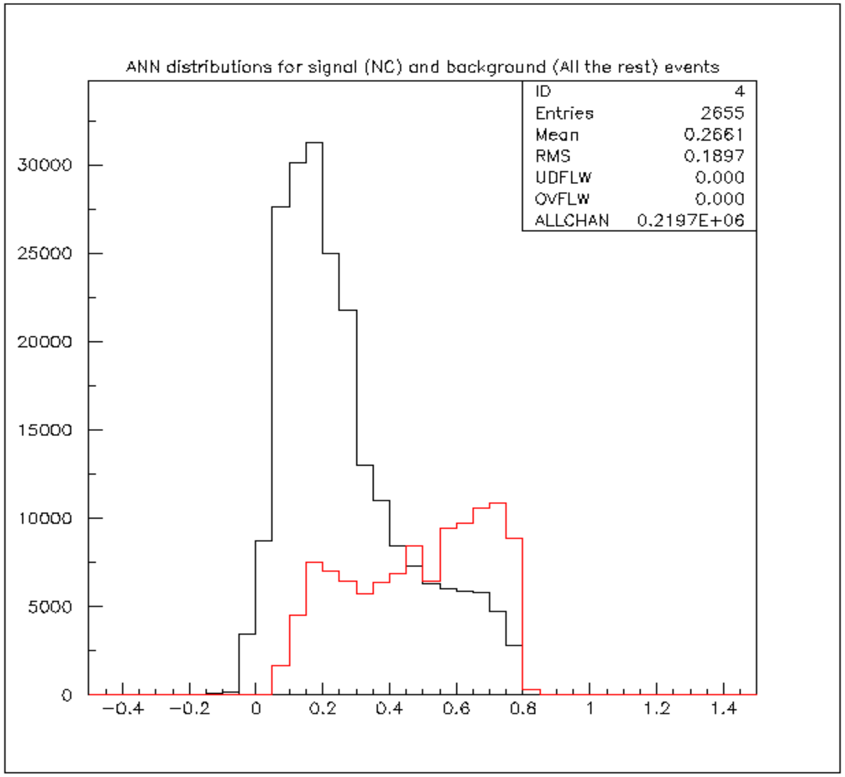


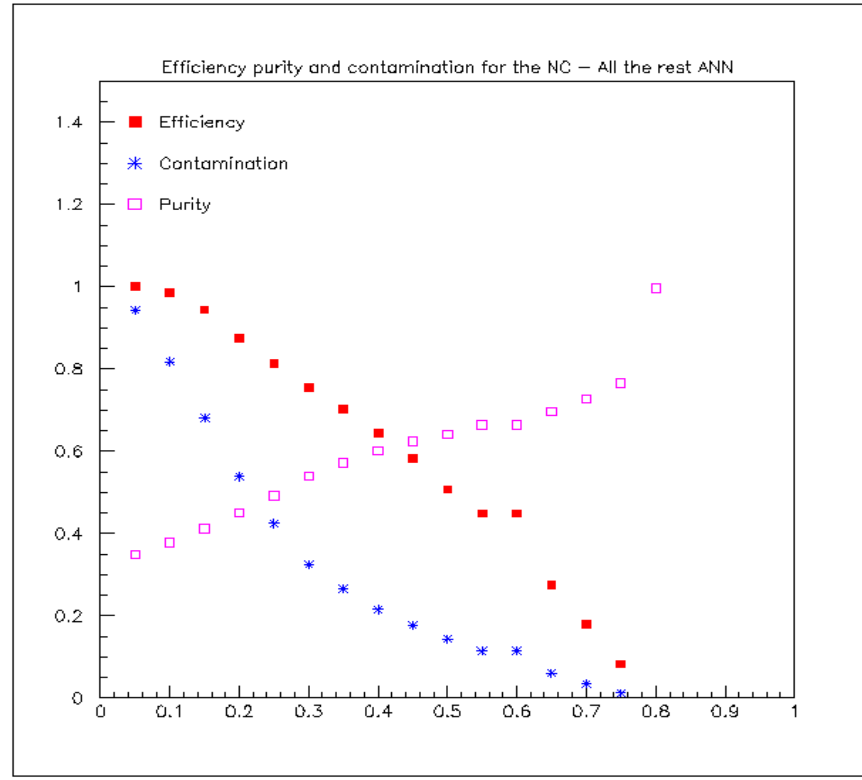


- If we consider a "cut" at 0.5 then we select signal (all events except ν_μ CC) with :

efficiency 97.2 % - purity 87.1 % - contamination 28.6 %

Results on NC - All the rest for period 4

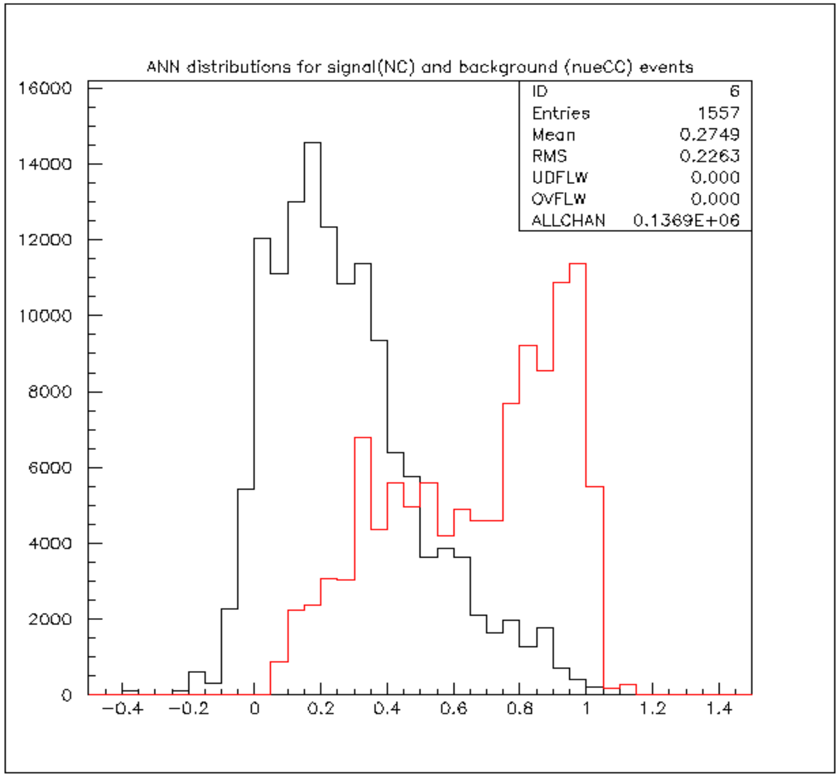


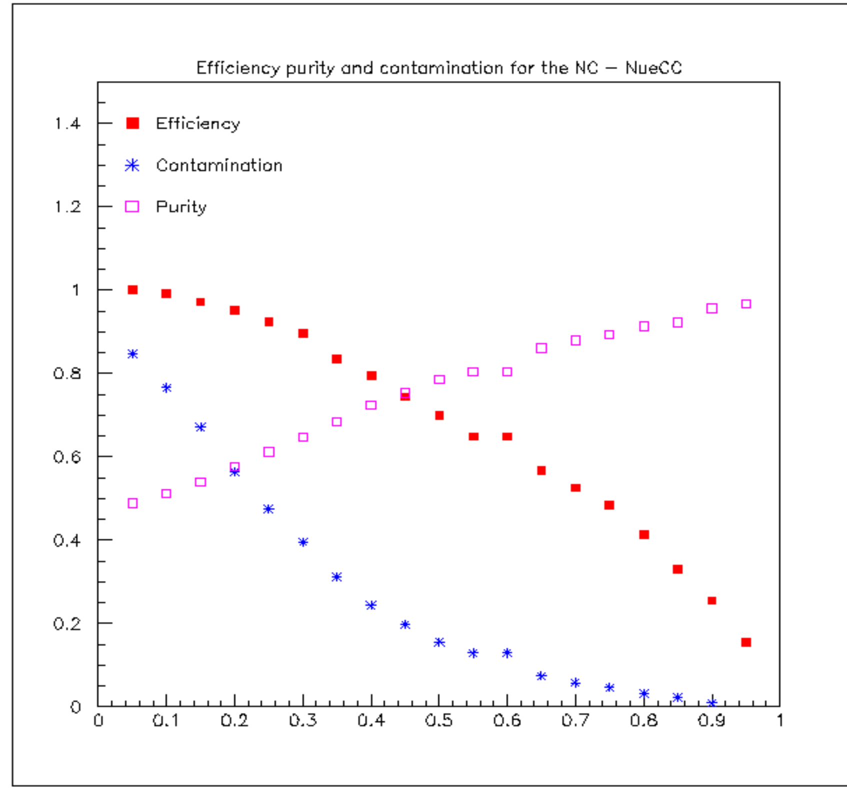


- If we consider a "cut" at 0.3 then we select signal (NC events) with :

efficiency 75.5 % - purity 53.9 % - contamination 32.5 %

Results on NC - ν_e CC for period 4





- If we consider a "cut" at 0.35 then we select signal (NC events) with :

efficiency 83.4 % - purity 68.4 % - contamination 31.2 %

Summary

- **@Efficiency ~ 87 %** (Good Statistics but relatively high contamination and poor purity)

PERIOD 3(4)

NETWORK	Efficiency	Contamination	Purity
ν_μ CC - All	87.1 (88.8)%	19.1 (19.8)%	89.1 (89.8)%
NC - All	88.1 (87.1)%	53.4 (53.8)%	46.9 (45.5)%
NC - ν_e CC	87.2 (89.5)%	33.3 (39.5)%	68.9 (69.7)%

- **@Contamination ~ 16 %** (Low Statistics but relatively high purity)

PERIOD 3(4)

NETWORK	Efficiency	Contamination	Purity
ν_μ CC - All	75.9 (80.1)%	14.5 (16.4)%	90.3 (90.7)%
NC - All	52.0 (50.7)%	16.3 (14.4)%	63.0 (64.0)%
NC - ν_e CC	74.9 (69.9)%	16.6 (15.5)%	79.2 (78.5)%

Conclusions

- All three neural networks show satisfactory results. The addition of new variables (?) could probably improve them further.

- The ANN for v_eCC - NC classification gives relatively better results than the set of ANN's for NC - All classification.

- All three neural networks do not use as input variables the hits in the SF system and the hits in the MID system since these do not show yet (at least for the SF system) great compatibility between data and MC.

- All neural networks will be checked with another (different) set of MC events to further test their performance.